heard during broadcasts were measured with the Monitor and assessed by the psychoacoustic test panel, showing good agreement. One of the questions needing further study is how will the frequency response of typical radio receivers alter the significance of the Loudness Monitor readings? Another is how well will the Monitor, connected to a sound level meter, indicate the loudness level of common sounds? With the introduction of the Loudness Level Monitor, an important step has been taken toward a correct and objective evaluation of the sensory loudness level of both the steady-state and impulsive sounds.

Acknowledgment

The authors acknowledge the continuous interest and support of Dr. P. C. Goldmark of CBS Laboratories, and of R. S. O'Brien, H. A. Chinn, J. D. Parker, O. L. Presthodt, J. L. Stern, and D. M. Vorhes of the Engineering staff of the CBS Broadcast Group.

References

[1] H. A. Chinn, D. K. Gannett, and R. M. Morris, "A new standard volume indicator and reference level," *Proc. IRE.*, vol. 28, pp. 1–17, January 1940.

[2] American Standards Institute (formerly American Standards Association), "Acoustical terminology," Standard S1.1, 1960.

[3] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement, and calculation," *J. Acoust. Soc. Am.*, vol. 5, pp. 82–108, October 1933.

[4] D. W. Robinson and R. S. Dadson, "A re-determination of the equal-loudness relations for pure tones," *Brit. J. Appl. Phys.*, vol. 7, pp. 166–181, 1956.

[5] American Standards Institute, "General purpose sound level meters," Standard S1.4, 1961.

[6] American Standards Institute, "Computation of the loudness of noise," Proposed Standard S3.4.

[7] B. B. Bauer and E. L. Torick, "Researches in loudness measurement," *IEEE Trans. Audio and Electroacoustics*, vol. AU-14, pp. 141–151, September 1966.

[8] S. S. Stevens, "Procedure for calculating loudness: Mark VI," *J. Acoust. Soc. Am.*, vol. 33, pp. 1577–1585, November 1961.

[9] K. D. Kryter and K. S. Pearson, "Some effects of spectral content and perceived noise level," *J. Acoust. Soc. Am.*, vol. 35, pp. 866–883, June 1963.

[10] D. W. Robinson and L. S. Whittle, "The loudness of octave bands of noise," *Acoustica*, vol. 14, pp. 24–35, 1964.

[11] I. Pollack, "Loudness of periodically interrupted white noise," *J. Acoust. Soc. Am.*, vol. 30, pp. 181–185, 1958.

The Transient Distortion of Loudspeakers and Its Evaluation

TOMÁŠ SALAVA

***Abstract*—The paper first briefly explains the causes of a considerable distortion of transient signals by electroacoustic transducers and reviews the methods used to investigate this type of distortion. Also discussed are certain questions of perceptibility and the importance of the transient distortion. There follows a brief description of preliminary listening tests with piano tones, carried out in order to examine certain disputable statements. Preliminary conclusions are drawn for the evaluation of transient distortion from the point of the perception of transient signals in music and speech.**

I. Introduction

ELECTROACOUSTIC transducers produce a relatively high amount of transient distortion, resulting mainly from the vibrations of the active mechanical elements. The highest values of this type of distortion occur in loudspeakers, especially in the conventional cone types. A distortion of transient signals originates in this case primarily as a consequence of the diaphragm's self resonances. Because of the relatively low damping of the vibrations and, at the same time, the relatively high density of the modes, particularly in the range of the medium and higher audio frequencies, there is produced a considerable distortion of transient signals, evident on an oscillograph as a change of the wave shape.

It is obviously possible to reduce, by suitable design, the influence of the vibrations of the diaphragm to a minimum.

Under certain conditions, however, it is more expedient to make a suitable use of the diaphragm's self resonances. Utilizing them, it is possible, for instance, to obtain an increased frequency range of the loudspeaker, or even a desirable modification of the frequency response, etc. In view of this, the diaphragm's self resonances should not always be reduced to an absolute minimum; it is preferable, in most instances, to reduce them to a value suited to each particular case. To this purpose it is, however, necessary to work out the corresponding criteria and methods of measurement and assessment of the above type of distortion. In view of this, an increasing attention has been paid lately to the problems of distortion of transient signals by electro-

Received June 8, 1967; revised September 6, 1967.
The author is with the Popov Research Institute of Radio Communications, Prague, Czechoslovakia.

acoustic transducers, both from the viewpoint of the physical function of the transducer and from that of its influence upon the perception of the sound quality.

II. A Review of the Methods Used in the Investigation of the Transient Distortion

As has been stated above, an optically observed distortion of transient signals originating in electroacoustic transducers may be of a considerable magnitude. The studies concerned with this fact date as far back as 1937.[1],[2] Of especial importance is the study undertaken by Shorter,[3] which was published in 1946. Shorter uses for his investigations an interrupted sine-wave signal, and he investigates the decay of the response at moments when the exciting signal has been interrupted. For evaluation of the transient distortion Shorter proposed a continuous recording depending on the frequency of the exciting signal. The values of the decay amplitudes are recorded at certain intervals after an interruption of the exciting signal (for instance, after 5 ms, 10 ms, 15 ms, etc.). The transient distortion is then characterized by a usual frequency response supplemented with a family of the decay characteristics. For this way of expressing transient distortion there is proposed and described a practical method of measurement. So far, Shorter does not take into account to any particular extent an evaluation of the transient distortion from the point of view of influence on the quality of sound.

As regards the physical function of a loudspeaker, the method proposed by Shorter suffers from one basic deficiency in that the moment of interruption of the exciting signal is not sufficiently defined. Moreover, with the exception of outstanding simple resonance the course of the decay is irregular, and a record of the amplitude only at certain intervals is also considerably doubtful. The shortcoming thus lies in the fact that the method of measurement does not clearly describe the decay phase of the response and is based on a "manual" compensation of the decay on a screen (the auxiliary voltage for a recorder being then derived from this compensation). Consequently, a further attempt was made to describe the transmission properties of the loudspeakers for transient signals in other ways, such as supplementing the frequency response of the acoustic pressure amplitude with an additional phase characteristic or with the group-delay characteristic.[4],[5] These attempts are based on a theory of circuits, elaborated primarily for circuits with concentrated elements. Such a description of the transient properties of loudspeakers has, however, only a limited applicability.

In 1950, Corrington[6] published a new method for measuring the transient distortion of loudspeakers, which so far has been the most widely used. An interrupted sine wave is again used as an exciting signal. The interruption, however, is performed regularly for a time which always corresponds to a definite and permanent number of cycles, the exciting signal always being switched on for the same number of cycles.

The switching-over itself is performed electronically in such a way as to make the instant of the switching-over action coincide with passage of the original sinusoidal signal through the zero value.

As a criterion of quality Corrington takes the mean value of the loudspeaker's response decay, that is, the mean value of the signal radiated by the loudspeaker at the time when the exciting signal is interrupted. The equipment used by Corrington is described by Kidd.[7] An exciting signal interrupted, for instance, after each 4 or 16 cycles is obtained with this equipment; and in the second part of the instrument, only those parts are keyed out of a signal coming from the measuring microphone that correspond to the decay at the times when the exciting signal is interrupted. The mean value of the decay is then recorded in dependence on the frequency under the conventional frequency response (of a steady state). A criterion of the quality is the difference of levels of the two recordings. This method has been adopted, for instance, by the Philips Company.[8]

Corrington also investigated the influence of transient distortion on the quality of sound reproduction and he stated[10] a very approximate conclusion, according to which any loudspeaker with less than 10 or 12 percent transient distortion (more exactly with an 18- to 20-dB difference between the steady-state frequency response level and the level of the mean value of the decay when the exciting signal is interrupted after each 16 cycles) will be comparable to the best available. However, the Corrington method has an importance primarily in the investigation of the physical function of loudspeakers, and for this purpose it was designed by the author in 1960.[9] But, instead of the originally proposed mean value of the decay, it is more advantageous to record the rms value which is more decisive, particularly where the courses of the decay are more complex. Where a record of the rms value is supplemented with an additional record of the peak value, it is possible to determine from the two recordings, for instance, the damping of the vibrations of the cone and the starting amplitude of the decay.

All of the above methods have been concerned so far only with the decay values of transducers. These are, however, in an overwhelming majority of cases, superimposed by reverberation. It is therefore necessary to trace the influence of transient distortion upon the quality of sound rather than on the signal buildup, which is much better preserved in listening rooms. In view of this, there has been designed and manufactured by the author an apparatus that makes it possible to record continuously the build-up distortion of an interrupted sinusoidal signal in the succession of individual cycles, this being realized continuously in dependence on frequency. By means of this apparatus it has been possible to obtain a series of curves, corresponding to the amplitude, by an output signal during the first, second, etc. cycle following the switching-on of the signal. From these curves it has been possible to reconstruct the course of the buildup of the resulting acoustic signal for any frequency. The smaller the build-up distortion is, in this case, the nearer all the curves are to the final record of a steady-state response.[9]

The latest of the more comprehensive papers dealing with the problems of transient distortion and offering relatively important conclusions[11] was published in 1961. The authors, Larson and Adducci, also base their considerations on the discovery that the decay is overshadowed by reverberation. They claim additionally that for familiar sounds, such as music, the ear will tolerate gross distortion of the wave envelope of a degree greater than those transients produced by the worst loudspeakers.

III. Experimental Part

For an examination of the statements by the two authors[11] dealing with the perceptibility of the transient distortion, a number of experiments were performed. The first was intended to verify the most important conclusion that had been taken over from Schaeffer.[12] He stated that, after as many as 50 ms have been removed from the start of a piano tone, the resulting alternation of the tone quality will still remain imperceptible. This, however, contradicts, for instance, the results of the studies[13] where a detailed analysis of piano tones is undertaken from the sound quality aspects.

For the performance of these experiments there were first recorded sequences of test tones played on a grand piano in the studio of the Czechoslovak Broadcasting Corporation. The recording was made by a small ribbon velocity microphone, Tesla-AMP-260. The frequency characteristic of the microphone used, with a high-tone correction, was ± 2 dB in a band of 50 Hz$-$18 kHz.

The recording was made at a velocity of 76 cm/s on a Telefunken portable studio tape recorder and played back on a 31/10 Philips tape recorder. The resulting playback frequency characteristics of the two units are within ± 1 dB in a band of 50 Hz$-$12 kHz. Used for reproduction were a Jensen G 610 Triaxial loudspeaker with a compression-type unit for the range from 400 Hz upward and a compression-type tweeter for the range from 7 kHz upward. This type of loudspeaker showed, relatively speaking, a very low transient distortion. The frequency characteristic measured in the window of an anechoic chamber, after a bass correction and an adjustment of the levels of the individual systems, was situated within ± 3 dB in a range of 60 Hz to 13 kHz. The resulting playback characteristic of the whole system was in the range of 80 Hz to 12 kHz within ± 3 dB. A level of the maximums at the beginnings of the strokes of 10 dB below the nominal input voltage of the tape recorder was maintained during the recording. During the playback, the maximum power input to a loudspeaker was one-tenth of the nominal input. Under these conditions, the harmonic distortion did not exceed the value of 2 percent over the entire frequency range.

A series of tones were recorded, each of them made up of 10 double strokes of the MF intensity, either single notes or quintaccords. For each second stroke the buildup was modified or the two strokes were left without any modification whatever. The listeners were asked to determine when the 2 tones of the pair sounded exactly alike and when there was a difference between them.

The build-up modifications were arranged in such a way that through a manual shifting of the tape there was first located the exact beginning of the stroke. From this beginning a length was then measured that corresponded to a set time interval; either the section within the marks was cut out exactly at right angles to the direction in which the tape moved and the tape was then joined again by means of the BASF adhesive recorder tape, or, alternately, the beginning was adjusted by a slanting cut, as shown in Fig. 1. The hatched portion was removed, the two parts of the tape were placed together and joined by the BASF adhesive tape. In this way, either a part of the recording up to a certain time interval from the beginning of the tone was removed or a larger or smaller modification of the tone or accord buildup was achieved. Fig. 2 shows an oscillographic record of the buildup of a standard musical pitch (440 Hz) produced by a stroke of an approximate MF intensity. Fig. 3 shows the oscillographic record of a buildup modified by a slanting cut of the tape a length of 100 m/s. For the listening tests a tape was made up with 10 series, each of which consisted of 9 pairs of strokes. Before the test was started, the listeners were instructed to concentrate upon the character of the beginning of the two tones, the first two series having been played on trial. The results of these first tests are shown in Fig. 4. The chart has been compiled according to data furnished from a total of 18 judgments for each case. From the figure it is first of all evident that a particular difference exists between the perceptibility of the "perpendicular cut" (so-called "tape editing") and the "slanting cut." A perpendicular cut up to 50 m/s is not perceptible at all in this case, while the slanting cut was perceived by listeners in 12 instances out of a total of 18. This is undoubtedly due to the fact that the buildup of a piano tone is itself very swift, so that the removal of a small portion of the buildup does not mean any appreciable change of perception of the tone quality.

In the first tests, the influence of a training or a sort of auditory experience on the part of the listeners was great. In illustration, Fig. 5 shows the results of a test performed with a well-trained listener and of another test with another listener who did not receive more than

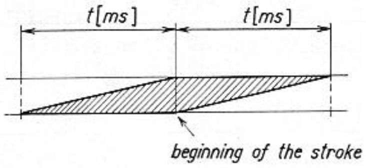

Fig. 1. Method of build-up modification by the slanting cut of the tape.

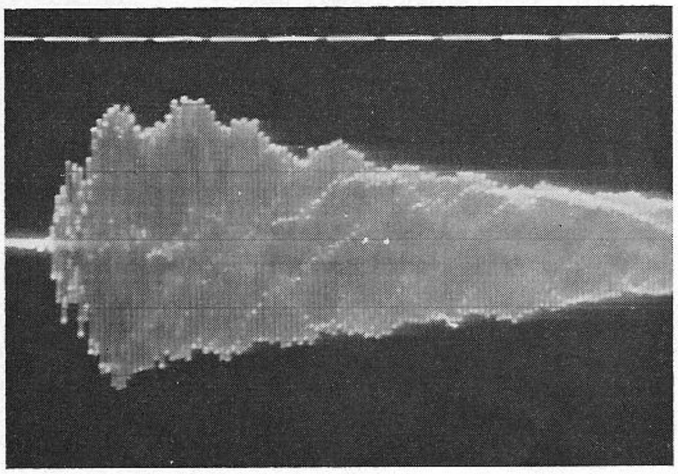

Fig. 2. The buildup of the standard musical pitch as struck on a concert grand piano—stroke "MF."

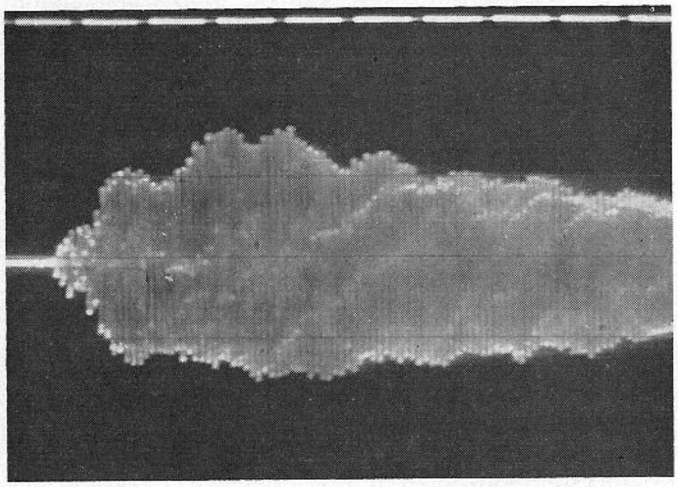

Fig. 3. The signal from Fig. 2, modified by the slanting cut of the build-up portion of the tape to a length of 100 ms.

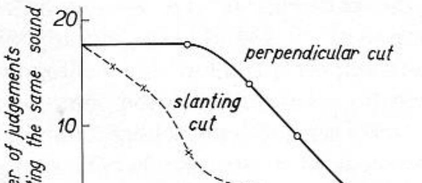

Fig. 4. Perceptibility of a modification through a perpendicular and a slanting cut.

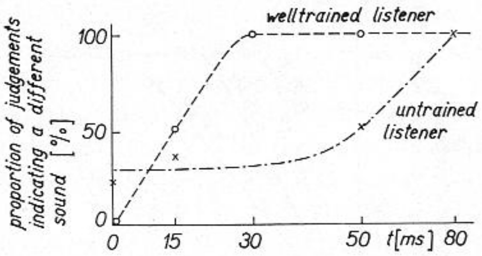

Fig. 5. Influence of the auditive experience of a listener on the perception of the build-up waveform.

a verbal instruction. For a definite evaluation 2 more tests with 10 trained listeners were performed.

The results for the c' 264 Hz, a' 445 Hz, and a'' 869 Hz are plotted in Fig. 6. They represent a value expressing the proportion of judgments indicating a different sound (different buildup) independence on the length of the cut. From the figure one clearly realizes a dependence of the perceptibility of a cut of an identical length on the frequency, as well as the fact that even a modification over a length as short as 15 ms already has a relatively high percentage of perceptibility.

In order to convey an idea of these values, let us attempt an approximation of a slanting cut by an exponential buildup. As an optimum, let us assume a time or duration of the cut equivalent to $1\frac{1}{2}$ time the reciprocal value of the build-up constant δ. With the length of cut in seconds defined as t, it is possible, under the given conditions, to regard the slanting cut as a sort of transient distortion produced by the resonance circuits, so that the build-up constant can be described thus:

$$\delta = \frac{2}{3}\,\frac{1}{t \cdot f}, \tag{1}$$

where δ is the build-up constant, t is length of the cut in seconds, and f is the frequency in hertz. Between the build-up constant or the damping constant of an output signal and the quality factor Q of the resonance circuit exists the following relationship:

$$Q = \frac{\pi}{\delta} = \pi\,\frac{3}{2}\,t \cdot f. \tag{2}$$

It is then possible to draw up, as an annex to the graphic representation in Fig. 6, a table of approximate equivalent values for the basic tone (Table I).

In order to be able to remove a distortion of the buildup produced by a slanting cut of the tape, it would be necessary to use a series of resonance circuits for the individual harmonics. The quality factor of these circuits would have to increase proportionately to frequency.

It is obvious that a modification of the buildup by a cut of the tape does not come too close to an authentic transient distortion of the buildup in loudspeakers, where the damping of the diaphragm's self resonances usually decreases only little with the frequency. How-

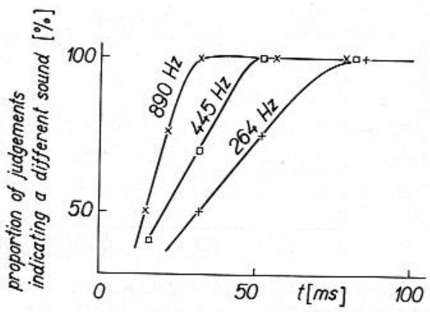

Fig. 6. Perceptibility of a slow-down of a piano tone buildup.

TABLE I

Perceptibility %	Quantity	Frequency of the basic tone		
		890 Hz	445 Hz	264 Hz
100	t ms	30	50	80
	δ equiv.	0.056	0.068	0.071
	Q equiv.	56	47	44
50	t ms	15	20	30
	δ equiv.	0.11	0.17	0.19
	Q equiv.	28.5	18.5	16.5

TABLE II

Q	500 Hz	1 kHz
7	20%	18%
15	35%	30%
20	55%	50%
30	80%	75%

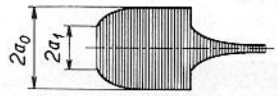

Fig. 7. Typical form of distortion of a test signal.

TABLE III

Q	a_1/a_0	500 Hz	1 kHz
30	0	80%	75%
	0.25	75%	75%
	0.5	65%	60%
	0.7	45%	50%

ever, it should be taken into consideration that the equivalent values Q of the resonance circuits, which we have obtained through the above approximation, are by no means high from the point of view of the values that are inherent to the vibration of the diaphragms. For a more exact verification, a simple experiment with an electric resonance circuit and an interrupted sine wave signal was carried out.

For listening tests a signal was used that was interrupted for an interval of 32 cycles and switched on again for another interval of 32 cycles. In the electric circuit it was possible to set up resonance frequencies of 500 Hz and 1 kHz, as well as Q from 0.5 to 30. The results of a preliminary experiment for a simple resonance, undertaken with trained listeners, are listed in Table II. During the tests, a distorted signal was compared with an equivalent undistorted signal. For reproduction a G 610 type loudspeaker was again used. For each frequency a total of 80 pairs (20 for each value of Q) were reproduced.

For transient distortion of loudspeakers it is, however, characteristic that the resulting signal is made up of a superposition of a distorted signal with an undistorted one and that it has the typical form illustrated in Fig. 7, where the starting amplitude a_1 and the steady-state amplitude a_0 are also marked out.

The same case can also be simulated by means of the experimental circuit used. Simple listening tests with this type of distortion were undertaken under the same conditions as the preceding ones. The results are listed in Table III, again as a percentage of correct judgments.

From the table it is evident that the perceptibility decreases much less than would correspond to a relative change of the form of the buildup. But, at the same time, the character of the sound changes, so that, for instance, the perceptibility of the instances $a_1/a_0=0$ and $a_1/a_0=0.5$ is higher than the average perceptibility between the two waveforms and the undistorted signal.

In these tests it was further discovered that sound character of the waveform according to Fig. 7 changes very markedly where any "fault" occurs in the buildup. So, for example, a sharp, very short (narrow) peak at the beginning of the signal results in a pronounced change.

The character of the sound changes markedly where the pulse amplitude in the buildup is only slightly higher than the starting amplitude a_1. If the buildup is entirely smooth, starting from a zero value, the sound character is "sinusoidally soft." This "softness" is lost, on the one hand, with a decreasing Q value (that is, with a decreasing distortion) or, on the other hand, with an increasing a_1/a_0 ratio. The acute pulse at the beginning produces an additional and pronounced increase of the so-called "loss of softness," to the point of a buzzing or even a cracking sound, periodically concurring with the periods of the switching of the test signal. A well-perceptible change of the sound arises already with a pulse that is nearly imperceptible in the oscillograph.

For illustration we may mention some of the results obtained in tests with speech, where a cutting-off of beginnings of words in a duration between 10 ms and 50 ms was effected.

This was, for instance, the case of the Czech word "Trám." When a section of 12 ms was successively cut off from the beginning, the word originally recorded was perceived as "krám," "prám," and, finally, as ám" (English equivalents of the words: "beam," hop," "ferry," "frame.")

The tests with words were undertaken on a larger ale. The results are not given in detail because in the rm in which they were performed they do not have y immediate relation to the problems under review. owever, they illustrate very well the complexity of rception of short signals, particularly at the beginngs of words.

IV. Conclusion

In electroacoustic transducers a considerable distoron of transient signals usually takes place. The princil reason for a distortion of transient signals in electrooustic transducers is the self resonance of the mechanlly active elements of the transducers.

As regards the physical function of a transducer, the orrington method of measurement may be taken as e most expedient, inasmuch as it registers the peak d the rms values of the decay part of the response. In is case it is possible to determine, from the results of e measurements, for each outstanding partial resonce, in addition to its resonance frequency, values ch as the degree of excitation and the equivalent facr of quality or damping. It is further possible to deterine by this method, for instance for a given frequency nge, the mean value of energy emitted into the oscilions, related, for instance, to the output radiated in a eady-state condition, etc.

Other methods of measurement, particularly the iginal Shorter's method and the methods derived om it, have a more indicative character, because the lues measured are not sufficiently defined.

As regards the perception of transient phenomena in tural acoustic signals, it has so far not been possible take an authoritative decision as to an optimum ethod of measurement. From practical experience it however, evident that an optical evaluation of the ansient distortion according to a change of the waverm, particularly by methods borrowed, for instance, om impulse technique, does not conform to the mechism of hearing and, consequently, leads to incorrect nclusions. In such instances, inasmuch as no checkup rough appropriate listening tests is performed, the ansient distortion by electroacoustic transducers is ually overestimated. As it is usually not possible to d a wider simple relationship between an optically sessed transient distortion and the current listening sts, in which the overall transmission properties of an ectroacoustic system are evaluated, there may, on the ntrary, arise some doubt as to the perceptibility of e distortion of transient signals in general. Such conusions are arrived at, for instance, by Larson and dducci. It is, however, necessary to agree with their atement that the influence of the distortion of trannt signals in electroacoustic transducers is usually overestimated in current commercial literature. It is also, undoubtedly, impossible to judge the quality of reproduction of transient signals according to two or three oscillographic records of a response to the interrupted-sine test signal, as through an appropriate choice of the frequency of the test signal it is usually possible, even in a loudspeaker of inferior quality, to find convenient forms of response and vice versa. (In spite of this, the method is frequently used in commercial literature.)

From the tests described above it appears that the auditive perception of even very short forms at the beginning of natural acoustic signals and a perception of the changes of these forms may be very good under certain conditions. Thus, for instance, a slow-down of the buildup of a piano tone by means of the so-called "slanting cut" of a tape over a length of 50 ms is quite well discerned by a listener without some experience in listening to piano music. When the buildup of a tone is modified by the "slanting cut" over a length of 30 ms, the perceptibility is still equivalent to 75 percent for experienced listeners. On the contrary, a simple removal of the beginning of a piano tone over a length up to 50 ms need not be perceptible at all. However, the removal of some 12 ms from the beginning of a word that begins, for instance, with a plosive consonant "t" may result in a distinct hearing of the consonant "k" instead of the original "t" at the beginning of the word. The above examples again illustrate, among other things, the fact that a simple change of the waveform of a transient signal need not lead to the simple conclusion as to what is the change in the quality of sound. The majority of studies that have investigated the distortion of transient signals in electroacoustic transducers have been based primarily on changes of the waveform of the investigated signals. It will probably be necessary to take as a point of departure a relative change of the spectrum of a signal while taking into consideration also the masking processes in the human ear. It is then possible, for instance, to explain satisfactorily the perception of very short impulses with a relatively low amplitude at the beginning of the investigated signals, etc. But not even this procedure can be dependably designated as fully meeting the present requirements.

The methods of measurement used for the investigation of transient distortion have to be considered rather from the point of view of a definition of the values measured and with regard to the ways in which the measurement results may be reproduced, that is, as a means of an objective evaluation of the physical function of a transducer.

Acknowledgment

The author wishes to acknowledge the valuable advice and comments received during his work from Prof. J. Merhaut.

References

[1] J. G. Helmbolld, "Oscillographische Untersuchungen bei Lautersprechern," *Akust. Z.*, p. 256, 1937.

[2] G. Backhaus, "Nichtstationäre Schallvorgänge," *Ergeb. exakt. Naturwissensch.*, p. 237, 1937.

[3] D. E. L. Shorter, "Loudspeaker transient response," *BBC Quarterly 1*, no. 3, 1946.

[4] W. R. Stroh, "Phase shift in loudspeakers," *IRE Trans. Audio*, vol. AU-7, pp. 120–124, September/October, 1959.

[5] C. A. Evaskio, "Electroacoustic phase shift in loudspeakers," *J. Acoust. Soc. Am.*, vol. 22, p. 444, 1950.

[6] M. S. Corrington, "Transient testing of loudspeakers," *Audio Engrg.*, p. 34, August 1950.

[7] M. C. Kidd, "Tone burst generator," *Electronics*, no. 7, p. 132, 1952.

[8] M. T. Haitjema, W. Kipinga, *et al.*, "Performance tests on loudspeakers," *Philips Tech. Rev.*, no. 1, pp. 23–33, 1960/1961.

[9] T. Salava, "Concerning the question of transient response measurement II," *Proc. 2nd Conf. on Acoustics* (Budapest), 1960.

[10] M. S. Corrington, "Correlation of transient measurements with listening tests," *J. Audio Engrg. Soc.*, vol. 3, no. 1, 1955.

[11] R. J. Larson and A. J. Adducci, "Transient distortion in loudspeakers," *IRE Trans. Audio*, vol. AU-9, pp. 79–85, May/June 1961.

[12] P. Schaeffer, "Note in time relationships," *Gravesano Rev.*, vol. 5, no. 17, 1960.

[13] E. Meyer and W. Lottermoser, "Über die Möglichkeit einer klanglichen Beurteilung von Flügeln," *Akust. Beihefte*, no. 1, p. 291, 1961.

Semiautomatic Speech Intelligibility Measurements

JOHN W. PREUSSE

***Abstract*—A recent study towards the improvement of intelligibility testing techniques has resulted in a very efficient testing procedure based on the use of the Fairbanks Rhyme Test and a computer. The Fairbanks Rhyme Test permits the actual tests to be taken quickly and simply, and the computer provides a rapid, thorough analysis of the test data. Subsequent evaluations of several vocoders, and of the effects of electrical stimuli upon test subjects, showed this testing procedure to be quite effective. Small differences in intelligibility between systems were detected by giving a large number of tests. A diagnostic evaluation of each system was performed by the computer, which pointed out specific shortcomings of each system under test.**

IT IS GENERALLY acknowledged that the only valid methods for evaluating the ability of a communication system to transmit intelligible speech involve the use of some form of speech intelligibility test. Unfortunately, even under the best of circumstances, intelligibility testing is a laborious, time-consuming task. Evaluations must be made, however, regardless of the difficulties, and, with the current growth of new ideas and developments in speech processing, the amount of testing which should be done is enormous. Therefore, it seemed reasonable to take a close look at current testing techniques in order to try to simplify and speed up the operation as much as possible. A study of this nature was begun about a year ago at the U. S. Army Electronics Command, Fort Monmouth, N. J. Effort was concentrated on the ways in which a digital computer could be used to automatically evaluate the test data, although other problems in testing were also considered. The study has resulted in the development of a testing procedure which provides a considerable time and labor savings in test administration and scoring, and permits a thorough analysis of the results. This procedure is termed semiautomatic because, although the subjects make their responses in the conventional way, the tests are scored and analyzed by a digital computer. A fully automatic system, in which the subjects would depress a key, or otherwise signal their response to an on-line computer, was not attempted due to the extreme complexity of such a system.

Considerable thoughtwas given to the problem of selection of suitable test material. A large number of well-known and not-so-well-known tests were considered for the application. The Fairbanks Rhyme Test[1] was ultimately chosen as the most suitable test in this particular case.

The Fairbanks Rhyme Test consists of 250 words arranged into 5 lists of 50 words each. Each word in a list shares a common ending with its counterpart in every other list. Thus, the first list contains the words hot, pay, top, etc., and the second list contains got, may, hop, etc. Answer sheets (Fig. 1) showing the word endings are supplied to the subjects. In each case, the first letter of the word is missing. The subject, upon hearing the word, attempts to write down the first letter in the space provided on the answer sheet. The percentage intelligibility of the system under test is the percentage of correct responses to a very large number of words that have been processed by the system under test.

The Fairbanks Rhyme Test has several important

Manuscript received June 19, 1967. This paper was presented at the 1967 IEEE International Conference on Communications, Minneapolis, Minn.

The author is with the U. S. Army Electronics Command Laboratories, Fort Monmouth, N. J.